Automatic Nursing Care Trainer Based on Machine Learning

Ankita Agrawal, Wolfgang Ertel
Institute for Artificial Intelligence
University of Applied Sciences Ravensburg-Weingarten
agrawala@hs-weingarten.de, ertel@hs-weingarten.de

Abstract

Nursing Care is a challenging occupation. The ergonomically correct execution of physically strenuous care activities is very important in order to avoid secondary health problems such as backache for the nursing staff. However, there is a scarcity of ergonomics experts to facilitate the education of caregivers. In the project ERTRAG (Virtual Ergonomics Trainer in the Nursing Care Education), we aim to develop a virtual trainer that supports learning of ergonomically correct movements, thus avoiding serious health risks. The virtual trainer itself is trained by means of machine learning techniques, while the virtual trainer observes a human expert. The project is funded by the German Federal Ministry of Education and Research.

1 Introduction

The need to deliver nursing care has increased over the recent years due to the challenges brought by the societal demographic changes and treatment advancements. Stagnating birth rates and continuously increasing life expectancy has led to long term changes in the age structure of Germany [Birg, 2003]. The hospital employees are confronted with growing physical strain in addition to the known mental stress. Furthermore, the increase in overweight patients is a major challenge for clinical professionals, which often leads to excessive demand. However, there is a lack of trained nursing staff in comparison to the increasing demand for health care services. Usually the nursing care students have a chance to attend seminars from the experts only two or three times during their entire apprenticeship. While taking care of the patients and elderly people, their health is at a constant risk. The caregivers often suffer with wok-related musculoskeletal disorders (MSD) [Serranheira *et al.*, 2014], especially back disorders and shoulder-arm complaints as they have to transfer heavy loads when working with patients. This partly results in significant occupational impairments and the loss of quality of life [Engels *et al.*, 1996; Kusma *et al.*, 2015; Freitag *et al.*, 2013]. Hence, the employees either go into premature retirement due to unfavorable working conditions and prolonged illness or have to take frequent sick leaves [Meyer and Meschede, 2016; Grobe, 2014], thereby increasing the urgent need for trained personnel. Here the virtual trainer supports the caregivers in the learning of ergonomically correct practices. It is also suitable for training the care-taking of a patient by family members at home. The system can be used to practice the basic care movements with a Kinect camera at home without straining the back muscles.

2 Problem Definition

In the project ERTRAG (Virtueller ERgonomieTRainer in der PflegeAusbildunG / Virtual Ergonomics Trainer in the Nursing Care Education), our goal is to develop a training system for the students and employees in the nursing profession that assists them with the training of basic daily care activities. The activities performed by students are recorded using cameras and shoe soles. A skeleton model is generated using the point clouds delivered by the cameras. Sensors attached to the shoe soles are used to measure the force carried by a person to find if the caregiver is lifting a heavy load. Machine learning is applied on the skeleton and force data to recognize the correct execution of an activity. Later while practicing the nursing care activities in front of the cameras, the error stances will be detected by the learned algorithm and an immediate real-time feedback in the form of audio messages, visual animation or through haptic sensors will be provided to the students. Possible individual improvements will be suggested or the expert video will be shown depending upon the severity and frequency of a particular mistake. In this way, the system will not only help maintain the working ability of older employees, but also in gaining young and skilled workers, thereby contributing to improving the quality and performance of a hospital. The project involves two research institutes and two companies from Baden-Württemberg, namely, University of Applied Sciences Ravensburg-Weingarten, University of Konstanz, TWT GmbH Science & Innovation and Sarissa GmbH, that bring in different areas of expertise to the system.

To get an overview of the various care activities and problems associated with the non-ergonomic movements, the first step is to consult kinesthetic and physiotherapy experts. After consulting experts and observing students in the skills lab, it became apparent that there is no standard movement sequence for performing an activity. The nursing staff adapts the movements depending on the factors such as weight of the patient, the kind of health problem and treatment prescribed to the patient. However, there are certain incorrect postures that should be avoided by the caregivers so as to maintain their health. Therefore, we dropped our earlier premise of recognizing one correct movement sequence and rather apply machine learning to classify the movements into correct ones and various error categories that should be avoided in any case. This makes the problem more challenging because an incorrect movement for a tall person may not be necessarily wrong for a small person. Also, it is not harmful if the back of a caregiver is bent normally as opposed to when the person is lifting a patient with the back bent in a wrong way. The classification task is described in detail in Section 3.3.

3 Technical Approach

For training the machine learning algorithm, a large labeled dataset is required. State of the art datasets for pose, activity and gesture recognition are publicly available. Some of the datasets are MSR Action 3D Dataset [Li *et al.*, 2010], MSR Daily Activity 3D Dataset [Wang *et al.*, 2012], MSR Gesture 3D Dataset [Kurakin *et al.*, 2012]. These datasets are available for specific tasks and actions such as day-to-day tasks involving brushing teeth, chopping vegetables, hand gestures, playing badminton, working on a computer and other human activities. However, due to the specific nature of the human posture data required by the care activities along with the shoe soles data, these datasets are not suitable for the ERTRAG system arising the need for our own data generation. The dataset should be comprised of the correct motion sequences along with the motion sequences containing incorrect stances of the caregiver for the three scenarios mentioned in Section 3.1.

3.1 Experiment Setup

In the project we observe three basic caregiving activities that are performed by the nursing staff. The frequently performed scenarios in a care facility are, (a) Moving a patient up in the bed towards the head as they often slide down in the bed, (b) Bringing a patient from the lying position in the bed to sitting position on the edge of the bed, (c) Transferring the patient from sitting position on bed edge to the wheelchair and vice-versa. In the first batch of data acquisition in 2017, the scenarios performed by a kinesthetic expert and two students are recorded using Microsoft Kinect v2 as shown in Figure 1. The second batch of data is currently being recorded with the help of a kinesthetic expert and about ten nursing students in different semesters. The students playing the role of patients are selected having different height, weight, gender so as to obtain a diverse dataset for applying machine learning. A wheeled hospital bed with the ability to elevate head/feet and adjust the bed height along with a wheelchair are also arranged for recording the nursing care activities in order to procure a genuine database for the problem scenarios. The movements are recorded in two hour sessions with 50 videos recorded for the three activities per session.

3.2 Dataset

The data was recorded with the help of an acquisition tool built using the API (Application Programming Interface) provided by the Kinect SDK (Software Development Kit). The recorded sample images for the scenario in which the expert transfers the patient from wheelchair to the bed are shown in Figure 2. The tool captures the RGB images, depth images, skeleton images and skeleton joint data for each scenario performed by the expert/students at the frame rate of 22 frames per second made available by Kinect. The skeleton joint data at each frame consists of the three-dimensional absolute position with respect to the camera and orientation in the form of quaternion for each joint. The tool can also be used to convert the image frames of a particular recording into a video sequence.

Figure 1: Setup for data acquisition with single-view camera with the adjustable bed and wheelchair.

For each activity, about 20 videos are recorded making it a total of 60 videos. The recorded data is then prepared for labeling. Performing one scenario takes on an average about 20 to 30 seconds. One RGB image per second is extracted from the recorded data using a python script. In total, there are 1454 images and 60 videos that have to be labeled.

To facilitate the data labeling by the experts and remove the need for local software installation, the author developed a web-based user-friendly labeling tool using the Google Web Toolkit as shown in Figure 3.

The tool is developed to label images and videos by the experts. The comparison of the labeling of images and videos will show whether static image data is adequate for the posture assessment or dynamic video data is essential. The tool takes an image or a video as input on the left side. The images are shown in a random order so that the data can be labeled based on the posture independent of the chronological order of the images in the execution of an activity. This ensures that the pose errors are accurately identified and the previous frames do not affect the labeling of a particular frame. Moreover, an error in the single frame does not make the whole sequence as incorrect but only the posture in this particular frame is assigned to be incorrect. If the image shows the wrong pose of the caregiver, the expert can assign an error category from the ones already available below the image and error severity in a range from 1 to 4. It is necessary to assign

Figure 2: The sample RGB, Depth and Skeleton images recorded with Data Acquisition Tool while the expert transfers the patient to bed.

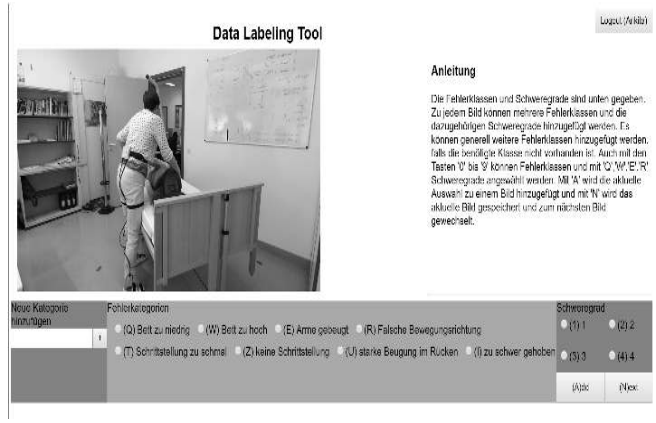

Figure 3: Web-based tool programmed for labeling the images and videos by the experts.

both error category and severity when an incorrect stance has been detected. If the desired error category is not available, a new category can be added that would be available for all the subsequent images and videos. If multiple errors in the pose of the caregiver are identified, multiple error categories along with their respective severity can be assigned to an image. If there is no mistake in the posture of the caregiver, that is, the expert has assigned no error to an image, the label for that image is automatically set to "correct".

The error categories correspond only to the unergonomic postures of the caregiver. The relative motion of the patient is not taken into account in the current analysis. Similarly, for labeling a video, when a pose error is identified, the video is paused and one or multiple error categories and their severity is assigned to this particular frame in the video. All other frames are labeled as "correct". It can happen that the errors at a particular frame are a result of the movement performed in the previous frames. Therefore, a fixed number of frames before the error frame would have to be observed by the learning algorithm while processing an error frame. The labeling can be carried out using either a mouse or the keyboard depending upon choice of the person using the tool. The data is labeled by two kinesthetic and one physiotherapy expert. After the completion of labeling, the skeleton joint data corresponding to the time stamp of the RGB images that are extracted for labeling is assigned the respective labels, resulting in a labeled set of skeleton data.

3.3 Feature Engineering and Classification

Since labeling is done by the experts independently, many of the error categories provided by them are different. The final set of error categories to be considered in the project are determined in a meeting with the experts. Some of the categories are combined together and the irrelevant ones are removed. The data labeled with the rejected error categories are labeled as "correct". The categories that are combined are renamed appropriately and the data is relabeled accordingly. The final eight error categories are,

1. Bed too low
2. Bed too high

3. The arms are bent
4. Movement in the wrong direction (the apprentice does not face the correct way while performing a movement)
5. Stride position is too narrow
6. There is no stride position present
7. Strong bending of the spine (while lifting the patients, the back should not be bent)
8. Patient being too heavily lifted (includes the cases when the plenum region such as back of neck or back of knee is grasped).

These final categories are in accordance to the fundamental ergonomically incorrect postures defined in the health care profession [Weißert-Horn *et al.*, 2014].

The results shown in this paper are obtained using the skeleton data recorded from Kinect to finalize the pose/motion analysis strategy. Kinect provides data for 25 joints, namely, SpineBase, SpineMid, Neck, Head, ShoulderLeft, ElbowLeft, WristLeft, HandLeft, ShoulderRight, ElbowRight, WristRight, HandRight, HipLeft, KneeLeft, AnkleLeft, FootLeft, HipRight, KneeRight, AnkleRight, FootRight, SpineShoulder, HandTipLeft, ThumbLeft, HandTipRight and ThumbRight. With the data acquisition tool, the absolute position and orientation in the form of quaternion provided for each joint at each time stamp is saved. Since the absolute position of a joint can vary for the same pose depending upon the position of the camera, relative coordinates of each joint with respect to the joint SpineBase along with their orientation quaternion are used as features. That is, the three-dimensional relative coordinates and four-dimensional orientation quaternion of all the joints at a particular time stamp forms one feature vector.

In the ERTRAG project we are dealing with the recognition of incorrect human postures while performing a nursing care task. Usually, skeleton or silhouette data is used for motion analysis and pose detection [Ye *et al.*, 2013; Elgammal and Lee, 2004]. However, due to the inherent task complexity, the classical methods of software problem solving are not applicable here. Therefore, supervised machine learning with automated feature generation to learn the different error classes is applied. After the labeled data captured from Kinect v2 has been obtained, this data is used to train different machine learning algorithms. The classification algorithms such as K-Means [Lloyd, 1982] variant for classification with k-means++ [Arthur and Vassilvitskii, 2007] initialization, k-Nearest Neighbors (kNN) [Cover and Hart, 1967], Support Vector Machines (SVM) [Cortes and Vapnik, 1995] and Extreme Gradient Boosting (XGBoost) [Chen *et al.*, 2015] are implemented and evaluated.

Pertaining to small amount of data and also to ascertain if the static data is sufficient, we first apply the algorithms as binary classifier. The positive data or the correct class ($label = 1$) consists of the data that has been labeled "correct" in the labeling tool. All the data containing non-ergonomic postures that are being assigned any of the error categories form the negative data and belong to incorrect class ($label = 0$). If the results prove to be good enough, the error categories will be used as individual labels to further train a multi-class classifier, otherwise the dynamic data or the movement sequences will be used. The skeleton data is normalized using Standardization technique. It normalizes the features by subtracting the feature mean and scaling to unit variance. The data is then randomly divided into 67% training and 33% test data containing feature vectors from both classes. The algorithm is trained on the training data using cross-validation [Kohavi, 1995] over a range of respective parameter values for each algorithm. For K-Means, the number of clusters is chosen between 2 and 9 representing the total available classes and k-means++ is used for initial cluster center calculation. The parameter ranges for kNN are:

- Number of neighbors - 1 to 26
- Weight function for prediction - Uniform, Distance

The parameters for SVM are varied as follows:

- Kernel - Linear, RBF, Polynomial
- Penalty term, C - between -2 and 10
- Kernel coefficient, gamma - between -9 and 3

The following parameter ranges are used for XGBoost:

- Number of estimators - 2 to 140
- Maximum tree depth - 2 to 6
- Learning rate - 0.05 to 0.8
- Minimum loss reduction, gamma - 0 to 10
- L1 regularization term, alpha - 0 to 50
- Minimum sum of weights of all observations - 0 to 50

The model with the best parameter combination is saved for each classifier. The learned models are applied on the test skeleton data to evaluate their performance and find the best fitting algorithm for the pose detection problem. Finally, the learned model of the best classifier will be used for real-time recognition of the incorrect movements.

4 Results

In this section, the results obtained for various machine learning algorithms on the labeling done by individual experts are discussed. Figure 4 shows the mean classification accuracy for the binary classifiers for the labels obtained from the two kinesthetic experts. As we can see, SVM performs fairly equally on both experts labeling with $80 \pm 3\%$ and $83 \pm 4\%$ accuracy, however, performs better with a mean accuracy of $90 \pm 3\%$ when the labels of the two experts are mixed (a feature vector is labeled as positive data and belongs to the correct class only if both the experts have not found any error in the corresponding RGB image). This is because in the beginning, the experts used different error categories to label the data. One expert focused on certain type of errors while the other expert assigned error categories such that some of them were slightly different. Therefore, the annotated data from both kinesthetic experts taken together yield improved results. XGBoost and kNN both give better results when the labels are mixed with $90 \pm 2\%$ and $88 \pm 2\%$ accuracy respectively. K-Means classification results are not shown as it performs very poorly with a mean accuracy below 35%. In general, we can see that the classifiers work better on Expert 2 labels which indicates that the labels assigned by Expert 1 are slightly inconsistent. Here we can also see that the classification accuracy does not vary significantly for SVM, kNN and XGBoost.

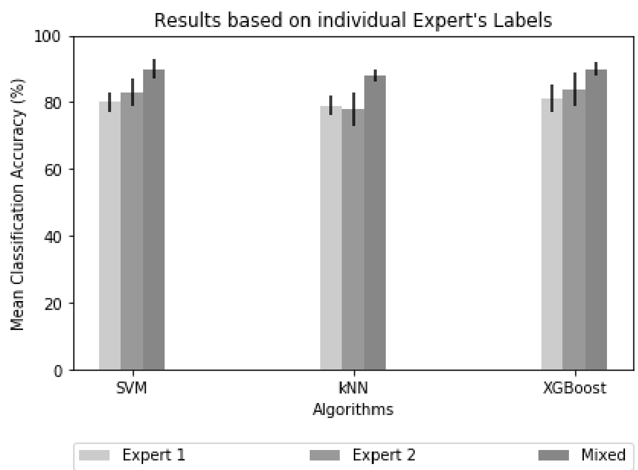

Figure 4: Results for labeling done by individual experts. Mean classification accuracy with lower and upper bound accuracy in percent.

The confusion matrix with and without normalization for XGBoost with mixed labels is shown in Figure 5 and Figure 6 respectively. In the figures, "correctPose" is the positive class and the "incorrectPose" represents the error classes. Out of the 480 test data, 414 data points are classified correctly as depicted in the diagonal elements. The off-diagonal elements represent the 66 data points that were misclassified.

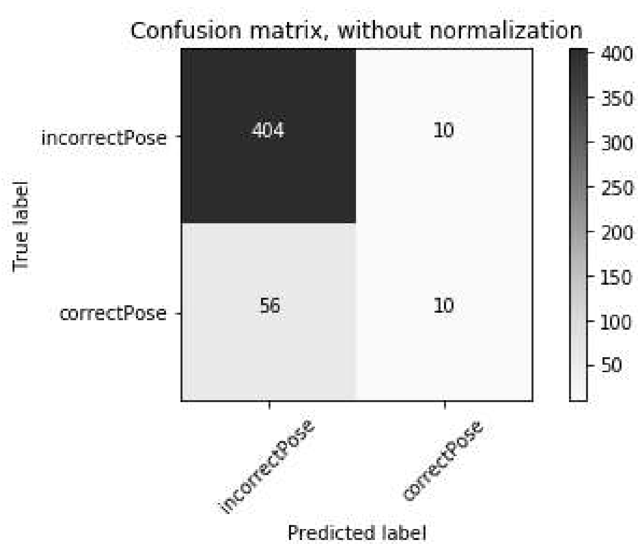

Figure 5: Confusion matrix without normalization on mixed data for XGBoost.

To evaluate the current performance of the classifiers on multiple classes, we executed them on the data with eight error categories and one correct category as mentioned in Section 3.3. The mean classification accuracy for the algorithms are shown in Table 1. The results are not good as we already expected but the renewed evaluation in coming months with a much larger dataset should give better results. The confusion matrix for the same is shown in Figure 7 and Figure 8. The error classes E1 to E8 correspond to the final eight error categories. The data contains no label corresponding to the error category "Bed too high". Therefore, E2 is not present in the confusion matrix. We can also see in the normalized confusion matrix that data belonging to E7 is mislabeled as E6, no stride position present. This may be because a data point labeled as E6 is often labeled as E7 as well by the experts.

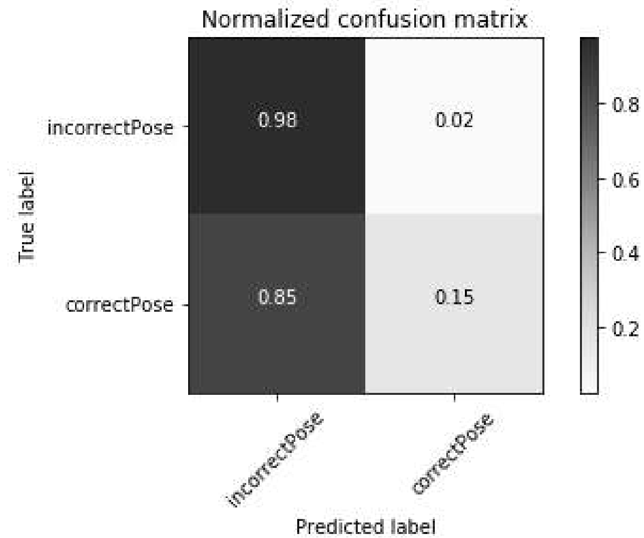

Figure 6: Normalized confusion matrix on mixed data for XGBoost.

Table 1: Mean Classification Accuracy (%) on Multi-class Classifier

Classifiers			
SVM	**K-Means**	**kNN**	**XGBoost**
68 ± 4	4 ± 0	67 ± 3	68 ± 5

5 Conclusion and Future Work

As can be seen in the results, SVM, XGBoost and kNN binary classifiers perform well on the static skeleton data producing $90 \pm 3\%$, $90 \pm 2\%$ and $88 \pm 2\%$ classification accuracy, respectively. The results also show that the multi-class classifier does not work very well as compared to the binary classification. However, it shows that the approach to use the static data should work and using a much larger database should improve results. If the binary classifier would not have given satisfactory results, it would be unlikely that the multi-class classifier would provide similar or better results. In that case, we would switch to the dy-

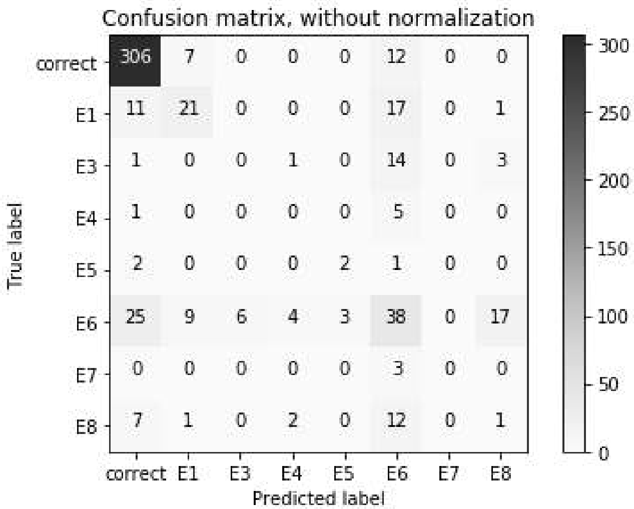

Figure 7: Confusion matrix without normalization on mixed data for XGBoost.

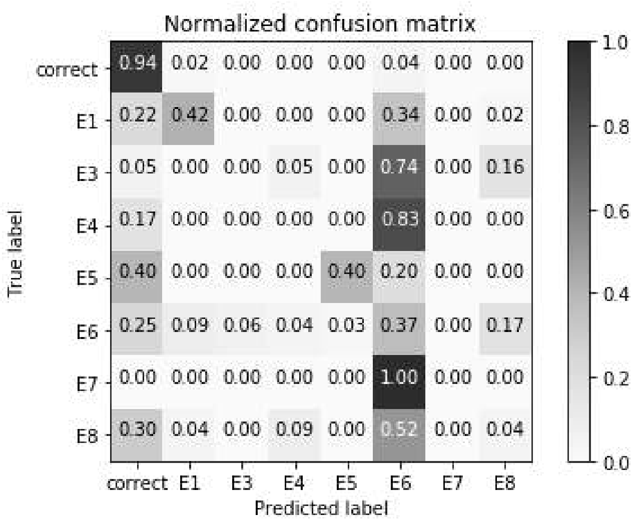

Figure 8: Normalized confusion matrix on mixed data for XGBoost.

namic data approach which involves observing the time series and applying relevant machine learning algorithms such as Markov Model [Lee and Nevatia, 2009; Lv and Nevatia, 2006] and Recurrent Neural Networks [Du *et al.*, 2015; Gers *et al.*, 1999] to find the incorrect postures and movements. Furthermore, in addition to the current setup where the training and test samples contain data from all the demonstrators, another setup would be analyzed. The second setup will leave one demonstrator out from the training samples and will only be used as test data so that this test subject has not been seen previously by the machine learning algorithm.

As already mentioned in the paper, a large dataset is favorable for obtaining better results. Currently we are collecting and labeling more data and we plan to optimize the current algorithms and evaluate the results. The recording is carried out using two cameras and force-measuring shoe soles. A regression algorithm will be applied to predict the error severity in addition to the error class. Other features such as Euler angles depending upon the degree of freedom of each joint will also be evaluated. If necessary, the dynamic data would be taken into account and machine learning would be applied to obtain better results. We will perform field tests in a health care institute to test the system. The feedback will be collected from the participating nursing care students and the results will be used to further improve our virtual ergonomics trainer.

References

[Arthur and Vassilvitskii, 2007] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[Birg, 2003] Herwig Birg. Dynamik der demographischen alterung, bevölkerungsschrumpfung und zuwanderung in deutschland: Prognosen und auswirkungen. *Aus Politik und Zeitgeschichte*, 53, 2003.

[Chen *et al.*, 2015] Tianqi Chen, Tong He, and Michael Benesty. Xgboost: extreme gradient boosting. *R package version 0.4-2*, pages 1–4, 2015.

[Cortes and Vapnik, 1995] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[Cover and Hart, 1967] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[Du *et al.*, 2015] Yong Du, Wei Wang, and Liang Wang. Hierarchical recurrent neural network for skeleton based action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1110–1118, 2015.

[Elgammal and Lee, 2004] Ahmed Elgammal and Chan-Su Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2004.

[Engels *et al.*, 1996] Josephine A Engels, JW Van Der Gulden, Theo F Senden, and Bep van't Hof. Work related risk factors for musculoskeletal complaints in the nursing profession: results of a questionnaire survey. *Occupational and environmental medicine*, 53(9):636–641, 1996.

[Freitag *et al.*, 2013] Sonja Freitag, Rachida Seddouki, Madeleine Dulon, Jan Felix Kersten, Tore J Larsson, and Albert Nienhaus. The effect of working position on trunk posture and exertion for routine nursing tasks: an experimental study. *Annals of occupational hygiene*, 58(3):317–325, 2013.

[Gers *et al.*, 1999] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. 1999.

[Grobe, 2014] T Grobe. Gesundheitsreport 2014. risiko rücken. *Gesundheitsreport-Veröffentlichung zum Betrieblichen Gesundheitsmanagement der TK, Hamburg*, 29(S 76), 2014.

[Kohavi, 1995] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.

[Kurakin *et al.*, 2012] Alexey Kurakin, Zhengyou Zhang, and Zicheng Liu. A real time system for dynamic hand gesture recognition with a depth sensor. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1975–1979. IEEE, 2012.

[Kusma *et al.*, 2015] B Kusma, J-J Glaesener, S Brandenburg, A Pietsch, K Fischer, J Schmidt, S Behl-Schön, and U Pohrt. Der pflege das kreuz stärken-individualprävention "Rücken" bei der berufsgenossenschaft für gesundheitsdienst und wohlfahrtspflege. *Trauma und Berufskrankheit*, 17(4):244–249, 2015.

[Lee and Nevatia, 2009] Mun Wai Lee and Ramakant Nevatia. Human pose tracking in monocular sequence using multilevel structured models. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):27–38, 2009.

[Li *et al.*, 2010] Wanqing Li, Zhengyou Zhang, and Zicheng Liu. Action recognition based on a bag of 3d points. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 9–14. IEEE, 2010.

[Lloyd, 1982] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[Lv and Nevatia, 2006] Fengjun Lv and Ramakant Nevatia. Recognition and segmentation of 3-d human action using hmm and multi-class adaboost. In *European conference on computer vision*, pages 359–372. Springer, 2006.

[Meyer and Meschede, 2016] Markus Meyer and M Meschede. Krankheitsbedingte fehlzeiten in der deutschen wirtschaft im jahr 2015. In *Fehlzeiten-Report 2016*, pages 251–454. Springer, 2016.

[Serranheira *et al.*, 2014] F Serranheira, A Sousa-Uva, and M Sousa-Uva. Importance of occupational hazards in nurses msd symptoms. In *Bridging Research and Good Practices towards Patients Welfare: Proceedings of the 4th International Conference on Healthcare Ergonomics and Patient Safety (HEPS), Taipei, Taiwan, 23-26 June 2014*, page 133. CRC Press, 2014.

[Wang *et al.*, 2012] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1290–1297. IEEE, 2012.

[Weißert-Horn *et al.*, 2014] Margit Weißert-Horn, Marianela Diaz Meyer, Michael Jacobs, Hartmut Stern, Heinz-Werner Raske, and Kurt Landau. "Ergonomisch richtige" arbeitsweise beim transfer von schwerstpflegebedürftigen 'Ergonomically correct'methods of transferring intensive care patients. *Zeitschrift für Arbeitswissenschaft*, 68(3):175–184, 2014.

[Ye *et al.*, 2013] Mao Ye, Qing Zhang, Liang Wang, Jiejie Zhu, Ruigang Yang, and Juergen Gall. A survey on human motion analysis from depth data. In *Time-of-flight and depth imaging. sensors, algorithms, and applications*, pages 149–187. Springer, 2013.